The manuscript is well-written, well-structured, and presents detailed investigation of mixed layer height (MLH) detection using mainly MicroPulse DIAL (MPD) lidar measurements, but also radiosondes, Doppler Wind Lidar and MRRR model. The authors have created nice and informative figures. They evaluate two retrieval approaches — one aerosol-based and one thermodynamically based — while also provide meaningful discussion on the different definitions of the planetary boundary layer (PBL), the diversity of retrieval techniques, and the challenges associated with their comparison. In my opinion, the work falls well within the scope of the journal and merits publication after some revisions.

This is a difficult topic, given the inherent challenge of comparing results from different instruments and retrieval algorithms. I find that the paper succeeds in providing well-supported assessment of MLH using multiple approaches with a focus on MPD observations. Nevertheless, there are several methodological aspects that in my opinion require further clarification to improve reproducibility and ensure the work can be applied in future studies.

Major comment:
The thermodynamic retrieval of MLH from lidar data using the parcel method is an interesting and novel contribution. However, as described in the manuscript, it appears to rely on several empirically set thresholds. The authors should elaborate further on the rationale for these thresholds and provide enough detail to allow the parcel method to be applied robustly to other lidar datasets.

Other comments/questions:

1. The abstract would benefit from specifying the locations where the measurements were conducted. This information is important for readers to interpret the environmental context of the results (for example it is not at the desert or a coastal site). Additionally, please clarify at the abstract whether the analysis was restricted to cloud-free condition.

2. Just a comment (no need for revisions): The introduction effectively emphasizes that, ultimately, the definition for the boundary layer is critical to ensure that comparisons make sense.

3. Lines 50–51: I suggest softening this statement. In real-world lidar measurements, it is common to observe a residual layer above the mixing layer. While the residual layer may not be part of the PBL in practice, it is often included in the conceptual model definition. This nuance is worth acknowledging to avoid oversimplifying the relationship between MLH and PBLH.

4. Lines 41-49: The authors give a solid overview of lidar-based PBLH retrievals and their challenges. Since your work deals with comparing and evaluating different methods and synergies, you might also briefly acknowledge recent efforts that use multi-sensor approaches (e.g., combining lidars, radiosondes, microwave radiometers, and/or models) to help reconcile differences in estimates from various techniques and synergies (Moreira et al., 2019, Tsikoudi et al., 2022, Chen et al.,

2022, and Zhang et al., 2025). Including a short sentence and these references would broaden the context for the study, like at the discussion (lines 497-499).

5. Line 54: Please provide a reference for the HRRR model here. You might consider repeating the Dowell et al. (2022) citation already used in line 203.

6. Section 2: Since aerosols are used as passive tracers in this study, I suggest adding a separate paragraph (or integrating into the existing text) discussing how they mix and exist in all the layers mention here. Specifically, maybe mention that aerosols are often trapped within the boundary layer and residual layers, and lidar measurements of aerosol backscatter allow us to detect these layers and define their tops. While this section is primarily theoretical and has been widely discussed in many previous studies, including this aerosol information could help readers better understand how lidar-based PBL detection works.

7. Fig. 1: Please define the abbreviation "AGL" as "above ground level". If I am not mistaken, this is the first occurrence of the abbreviation in the manuscript.

8. Lines 75–77: I suggest noting that the residual layer is detached from surface properties, and while some may argue it is not strictly part of the PBL, at the end of the day its inclusion depends on the definition used. It could also be useful to cite studies that have investigated the residual layer using lidar, for example Fochesatto et al., 2001.

9. Line 87: The term "transition layer" is generally difficult (if not impossible) to capture by lidars due to overlap effects, which is also the case for MPD measurements.

10. Lines 91–98: The discussion here makes perfect sense, but could the authors provide references to support these statements?

11. Line 103: From a quick search, it seems that Tonopah is surrounded by several notable mountain ranges and peaks. How can it be described as a "flat basin" given its elevation of 1641 m? Additionally, there is a statement regarding Tonopah's MLH—has this been investigated in other studies or it is an empirical statement? What are the corresponding climate characteristics at the Colorado site? The measurement period seems short and may represent only limited atmospheric conditions. Finally, it would be helpful to clarify the typical aerosol types encountered at these sites (e.g., biomass, pollen, dust), especially because they are used as tracers for the MLH. Both Boulder and Tonopah are located near notable terrain features (e.g., Cheyenne Mountain for Boulder; Lone Mountain and Butler Siebert for Tonopah).
    Have the authors considered the impact of complex terrain on their MLH observations? Are there

any previous studies investigating orography effects at these sites? Also, both sites are around 1600 m elevation, but their surface characteristics differ (forested vs. arid); could this influence the measurements or interpretation?

12. Line 115: HSRL measurements still require careful instrument calibration (e.g., molecular channel normalization, overlap correction near the surface) and quality control. It would also be helpful to mention the wavelength of operation.

13. Line 116: How is the virtual potential temperature calculated? Are dry and moist air considered, and are clouds excluded? Is the water vapor mixing ratio used? How is pressure incorporated? This information, along with the relevant formulas, should be described in an appendix. In my opinion, since this is the first time the parcel method is applied to lidar data, a detailed methodological description is necessary.

14. Lines 118–119: When the authors say "*the MPD had a minimum range of 318 m AGL and a vertical resolution of 150 m with 37.5 m range bin spacing,"* do they mean that the MPD averages 4 bins to improve the signal-to-noise ratio (4 × 37.5 m = 150 m)? So, the meaningful profile resolution is 150 m, while the raw data points are spaced every 37.5 m and are noisier? Maybe I am missing something, please clarify.

15. Line 125: I suggest, if possible, to provide references for the Halo Streamline Pro and Vaisala instruments. Additionally, were any corrections applied for noise? What type of scan was performed (horizontal, vertical, or both)?

16. Lines 130–131: Could the authors clarify why precipitation data are needed for the DWL? Does the instrument automatically stop or flag data during rain to avoid contamination, or is this used to filter MLH retrievals?

17. Line 148: How are cloudy data removed from the lidar observations? Does the algorithm apply a threshold on the raw signal? Removing clouds is critical in lidar studies, and it would be helpful to clarify how the method distinguishes clouds from dense aerosol layers, which can also produce strong backscatter. Lines 163 and 186–187 suggest that some cloud-affected measurements remain in the dataset; please elaborate on how these are treated or flagged.

18. Appendix A – Is b the translation parameter? Please clarify, it may be confusing for the reader. I understand that r_b and r_t correspond to the bottom and top of the integration, but the role of b should be explicitly described. How is b defined within the piecewise function?

19. Line 152: If you are referring to Dijkstra (1959), please cite it directly, even if it is already mentioned in de Bruine et al. (2017). This makes it easier for readers who encounter the algorithm for the first time. I think it is a pretty common algorithm in informatics, but not really in atmospheric physics. Also, could the authors clarify whether Pathfinder is a software package used for retrieving the MLH, or if the algorithm was implemented from scratch? Was it applied to 5-minute backscatter profiles? If yes (as mentioned in L53) was it attenuated backscatter? These details will help better understand the methodology.

20. Lines 156–159: The range for the top limiter seems quite large. Is it empirically set? Also, how does the algorithm handle cases where aerosol layers exist close to the top of the mixing layer, or when strong horizontal winds enhance turbulence and lead to unusually high MLH values?

21. Line 163: Again, it is unclear how low clouds are identified and removed from the lidar data. Are specific thresholds applied to the signal? How are these distinguished from dense aerosol layers that could produce similarly strong backscatter? A clear description is needed.

22. Line 176: Please clarify whether this weather station is the same as the one mentioned in line 130. If it is integrated into the MPD system, this should be also stated in Section 3.3.

23. Lines 178–187: It seems that many of these thresholds and limits were set empirically. A visual representation (e.g., a diagram or flowchart) would help the reader understand the workflow. This is especially useful for someone applying the thermodynamical parcel method to their own lidar data. For instance, a figure showing how the virtual potential temperature is calculated, including how different time resolutions for humidity and temperature are handled, would make the method more transparent. I would suggest, if not included in the main text, consider adding it to the appendix.

24. Line 193: The parcel method using a 1 K offset from the surface value, is applied in the same way as for the MPD? So that the initial condition or reference for both methods is consistent (or, if different, how it differs).

25. Lines 194–201: The comparison between radiosonde–Richardson and MPD–aerosol MLH retrievals is reasonable for daytime convective conditions, but it can diverge in transition periods or multi-layer situations. After sunset or in late morning, elevated well-mixed layers or residual aerosol layers can persist aloft, so the Richardson method (sensitive to turbulence) may detect a different top than the aerosol-based MLH. The authors partly mention this (L202–203), but it should be emphasized that their comparison is valid for daytime convective periods, as the algorithm does not run at night. It would also help to specify until what local time the algorithm provides retrievals. Is it 20:00 local time? Additionally, in my opinion it is worth expanding on the Richardson method in the appendix, similar to the treatment for Wavelet Haar and the wind lidar, to clarify its application.

26. Lines 220–221: It should be briefly clarified how virga was manually removed from the lidar data. For example, was a threshold applied to the attenuated backscatter or signal strength? Providing

this detail helps readers understand how aerosol layers were distinguished from falling precipitation or virga.

27. Lines 228–229: Could you clarify whether the parcel method was also applied to the HRRR virtual potential temperature fields? If so, this strengthens the need to include the full description and formulas of the parcel method in the appendix.

28. The authors have done a great job discussing the 6 and 8 September case studies. The manuscript clearly contrasts different boundary layer conditions and aerosol structures, effectively demonstrating how the methodology performs.

29. Section 5.2 is particularly valuable, as it compares the parcel method applied to different instruments—radiosonde (in-situ) and lidar (remote sensing). These instruments measure different tracers for MLH, making the comparison insightful, and the authors describe the results well. However, clarification on collocation is needed: Were the lidar products time-averaged around the radiosonde launch? For example, was the closest 5-min backscatter profile used for the MPD-aerosol method, 10-min water vapor, and 40-min temperature for the MPD-thermodynamic method?

30. L355: Do you mean that the MLH retrieved from the parcel method is on average 250 m higher than the MLH from the Richardson method? If so, why this difference occurs? Both methods detect turbulent layering, but the Richardson method responds directly to shear and turbulence, whereas the parcel method may follow the thermodynamic profile, potentially resulting in systematically higher MLH estimates.

31. L346-349: The description of the different comparisons in each column could be clarified. From the figure, it seems that (a) and (d) show all conditions, (b) and (e) show clear-sky conditions, and (c) and (f) show clear-sky conditions at 15:00 local time. I believe it would help the reader if this were explicitly stated in the text. Also, please clarify how clouds are defined or detected in your MPD retrievals, since this affects which points are included in the clear-sky subsets.

32. Fig 5: Consider maybe adding a small legend for the two lines at the bottom right to make it easier for the reader to distinguish them quickly. Also, using the same solid black line for both radiosonde retrievals (panels a, c) and MPD retrievals (panel b) can be confusing, even if explained in the caption.

33. Section 5.3: It seems that each point in Figure 6 represents roughly 2 hours, given that 0 corresponds to 06:00 LT and 1 to 20:00 LT (14 hours divided by 9 points). Is it true? It would be helpful to clarify the model's time step/resolution, and this information, along with the spatial resolution, should be included in Section 4.5.